DOCUMENT RESUME

ED 276 732 TM 850 617

AUTHOR Wainer, Howard
TITLE Estimating the Number of Examinees Who Did Not Reach the Last Item of a Section.
INSTITUTION Educational Testing Service, Princeton, NJ. Program Statistics Research Project.
REPORT NO ETS-RR-85-20; ETS-TR-85-60
PUB DATE May 85
NOTE 12p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Entrance Examinations; *Difficulty Level; Estimation (Mathematics); High Schools; *Item Analysis; *Response Style (Tests); Testing Problems; *Test Items; Test Length; *Timed Tests; Transformations (Mathematics)
IDENTIFIERS Scholastic Aptitude Test

ABSTRACT

It is important to estimate the number of examinees who reached a test item, because item difficulty is defined by the number who answered correctly divided by the number who reached the item. A new method is presented and compared to the previously used definition of three categories of response to an item: (1) answered; (2) omitted--a nonresponse followed, immediately or eventually, by a marked item in the same section; or (3) not reached--a nonresponse when no subsequent item in the section is answered. This practice makes it impossible to determine that the last item of a section was omitted, and may result in a biased (too easy) estimate of difficulty. The proposed method uses Tukey's Flog, a logistic transformation which enables linear extrapolation to plot the proportion of examinees who reach an item against the item number. Using 25 items from the Scholastic Aptitude Test as an example, 421 more examinees reached the last item than previously estimated, changing the item difficulty from 17 to 17.5. This proposed modest change appears to improve measurement accuracy, particularly when the estimated item difficulty must remain stable despite changes in item position. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED276732

RB-85-28

Estimating the Number of Examinees Who Did Not Reach the Last Item of a Section

Howard Wainer

PROGRAM STATISTICS RESEARCH

TECHNICAL REPORT NO. 85-28

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H.C. Weidenmiller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EDUCATIONAL TESTING SERVICE
Princeton, New Jersey 08541

TM 850 017

Estimating the Number of Examinees Who Did Not Reach the Last Item of a Section

Howard Wainer

Program Statistics Research
Technical Report No. 85-60

Research Report No. 85-20

May 1985

Copyright © 1985 by Educational Testing Service. All rights reserved.

The Program Statistics Research Technical Report Series is designed to make the working papers of the Research Statistics Group at Educational Testing Service generally available. The series consists of reports by the members of the Research Statistics Group as well as their external and visiting statistical consultants.

Reproduction of any portion of a Program Statistics Research Technical Report requires the written consent of the author(s).

ABSTRACT

A statistical method is proposed and illustrated for estimating the number of individuals that did not reach the last item on a section. This is compared with the current method and shown superior.

Table of Contents

		Page
I.	Introduction	1
II.	Estimation Method	2
III.	For Example	2
IV.	Conclusion	3
V.	References	3

Estimating the Number of Examinees Who Did Not Reach the Last Item of a Section

Howard Wainer
Educational Testing Service

I. Introduction

Current practice defines three categories of response to an item. These are:

1) Answered — that is, marked.

2) Omitted — a nonresponse followed, immediately or eventually, by an item in the same section that is marked.

3) Not Reached — a nonresponse when no subsequent item in that section is answered.

It follows from these definitions that it is impossible to OMIT the last item in a section, since one cannot mark a subsequent item.

This naturally leads to two questions.

i) Is this a sensible way to do it?

ii) Does it matter?

Dealing with the second question first, I believe that it is important to estimate the number who reached an item as well as we can. The difficulty of an item is determined by the number of examinees who answered the item correctly divided by the number who reached the item. If we make it impossible to OMIT the last item of a section, these items will tend to show fewer people reaching them then is probably the case. As a result, they will tend to look easier than would be the case if their location in the test was changed.

The current method, that of saying that all who did not answer the last question "did not reach it" is a lower bound on the number who reached it. We can often do better. This note proposes one way to do so.

II. Estimation Method

If we assume that the fall off of "reachers" occurs in a reasonably regular way, a method for estimating this emerges immediately. Let us consider just the last n items in the section (say the last 6). If we plot the proportion who reached each item against the item number we find that the scale is often so compressed as to make accurate extrapolation difficult. Thus, we shall first transform the proportion reaching each item into its associated logit. The transformation we shall use is a specific logistic transformation called (by Tukey, 1977) a Flog. Specifically, if p_i is the proportion reaching item i, then

$$f_i = \text{flog}(p_i) = 1/2 \ln(p_i/1-p_i) . \qquad (1)$$

This transformation stretches out the tails of the distribution in a way that makes linear extrapolation easy. One then fits a straight line to the last few points and estimates the value of f_i for the last item and back-transforms into the appropriate metric, i.e.

$$p_i = \exp(2f_i)/(1+\exp(2f_i)) . \qquad (2)$$

III. For Example

In Table 1 are data showing the 25 items in an SAT section, along with the number who reached each item, the proportion who answered it correctly (p+ = number correct/number reaching), and the flog of the proportion reaching each item.

Insert Table 1 and Figure 1 about here

In Figure 1 is shown a plot of the flogs of items 19 through 24 along with a linearly extrapolated value for the last item. Also shown is the flog of the proportion reaching under current definitions. The linearly extrapolated value seems more sensible. After back transforming the estimated flog (=1.09) into "number reaching" we have the results seen at the bottom of Table 1. We see that the number reached under this formulation is 2411 rather than the 1990 under current definitions. Note that if one uses 2411 as the number reaching, the difficulty of item 25 (delta) changes from 17 to 17.5; not an enormous change, but it is something.

IV. Conclusion

This proposal will effect only modest changes in current item analysis methodology, but if further investigations bear it out, it will add a little bit more accuracy to our results. This will be especially important in applications which require that the estimated item parameters (especially difficulty) remain relatively stable despite modest changes in item position. I believe that it would be worthwhile to try out this scheme in a variety of contexts and see if it continues to yield plausible results.

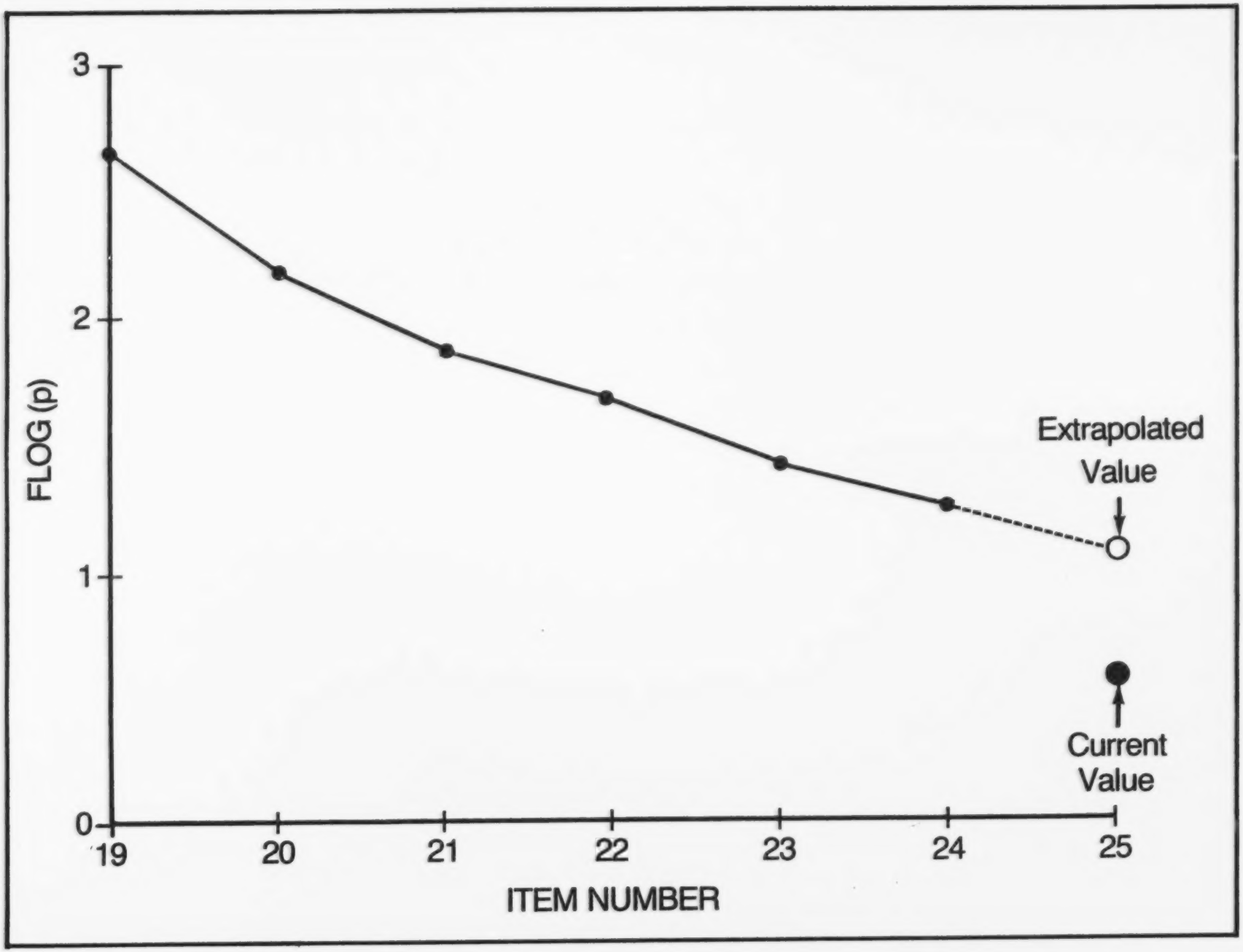
V. Reference

Tukey, J.W. (1977) Exploratory Data Analysis. Reading, MA: Addison-Wesley.

My thanks to Marilyn Wingersky for the data in the example.

Table 1

ITEM	NO. OMITS	NO. REACHED	PROPORTION REACHED (P)	PROPORTION CORRECT (P+)	FLOG (P)
1	5	2684	1.00	.94	
2	38	2684	1.00	.85	
3	70	2684	1.00	.83	
4	3	2684	1.00	.79	
5	61	2684	1.00	.79	
6	68	2684	1.00	.74	
7	106	2684	1.00	.75	
8	157	2684	1.00	.76	
9	199	2684	1.00	.68	
10	113	2684	1.00	.71	
11	95	2684	1.00	.60	
12	70	2684	1.00	.59	
13	61	2683	1.00	.59	3.95
14	298	2680	1.00	.56	3.25
15	107	2680	1.00	.49	3.25
16	356	2678	1.00	.41	3.05
17	255	2678	1.00	.40	3.05
18	53	2673	1.00	.74	2.75
19	128	2671	1.00	.08	2.66
20	286	2651	.99	.27	2.19
21	511	2623	.98	.22	1.88
22	264	2596	.97	.31	1.69
23	529	2539	.95	.43	1.43
24	48	2484	.93	.18	1.26
CURRENT 25	0	1990	.74	.16	.53
ESTIMATED 25	421	2411	.90	.13	1.09


3
2
1
0
FLOG (p)
Extrapolated
Value
Current
Value
19
20
21
22
23
24
25
ITEM NUMBER